The effect of attention on electrophysiological measures of syllable processing: contrastive features sorted according to a secondary contrast

Morris, David Jackson; Tøndering, John; Pharao, Nicolai

Published in:
Proceedings of the 19th International Congress of Phonetic Sciences

Publication date:
2019

THE EFFECT OF ATTENTION ON ELECTROPHYSIOLOGICAL MEASURES OF SYLLABLE PROCESSING: CONTRASTIVE FEATURES SORTED ACCORDING TO A SECONDARY CONTRAST

David Jackson Morris, John Tøndering Nicolai Pharao

Department of Nordic Studies and Linguistics, University of Copenhagen
dmorris@hum.ku.dk, johnt@hum.ku.dk, nicolaip@hum.ku.dk

ABSTRACT

While it is acknowledged that phonological processing begins at an early perceptual stage, the interaction of attention with contrastive speech features is inadequately understood. We investigated the effect of attention on electrophysiological responses to contrastive VOT, vowel length and place of articulation. Specifically, we examined the attentional difference in global field power for contrasts sorted according to secondary contrasts.

The attention-contrast difference timeseries that included VOT showed distinct differences between the two secondary levels, occurring less than 132 ms after the mean time at which the contrasts were differentiated in the stimuli. Moreover, there was temporal congruence between the initial point from which the two levels maximally differ for VOT and vowel length, but not for the vowel length-place of articulation contrast or its inverse.

These results elucidate the integration and automaticity involved in syllable processing, and they confirm that the cortical response to VOT is robust.

Keywords: Voice-onset time; Place of articulation; Vowel length; Electrophysiology; Global Field Power; Event-related potentials

1. INTRODUCTION

The effect of phonological features on neuronal measurements, particularly VOT, have often been studied with passive attention paradigms where the listener watches a silent film or reads a book. The results from these paradigms differ from attend conditions, where the amplitudes of event-related potential components are generally larger and the latencies shorter. The difference between attended and nonattended conditions, particularly when attention is more fully diverted from the auditory stimuli, allows the investigation of the attentional difference evoked by different features occurring in the stimuli. The present study probes these attentional differences in the context of three syllabic contrastive features which are VOT, place of articulation and vowel length. All three of these features are used to signal lexical contrast in Danish, which was the language environment in which testing took place.

The underlying premise of analyzing contrastive features of speech according to differences in the attentional state of test subjects is that it allows for disentangling of obligatory and attended processing associated with a contrast. This premise also attributes a central role to attention in speech contrast processing. Subtraction of conditions is warranted given that processing of contrastive vocalic content occurs even when attentional resources are not allocated to speech [1]. Also, the study of lexical effects has shown a degree of automaticity in phonological processing that is evident less than 200 ms after the auditory presentation of a speech feature [2].

Our analysis is based on the temporal attributes derived from Global Field Power (GFP) [3] timeseries. GFPs are a suitable measure for investigating biomarkers of contrastive features, as they reflect the standard deviation of the mean activity from all electrodes. Due to this, they are reference-independent and superior to single sensor average event-related measurements as they capture a broader range of scalp activity in the timeseries, not just that associated with single or clustered recording locations. We examine the difference of the attention conditions on secondary contrasts, that is, when the neural timeseries from a contrastive feature is sorted according to another secondary contrast. Our analysis focuses on the correspondence between auditory presentation and the observed changes in the group GFP attention-contrast differences. We attribute these changes to differences in the processing of contrastive features.

2. METHOD

2.1. Subjects

EEG was recorded from 20 University staff and students (10 identified as female; mean age 25 years) all of whom reported no existing neurological condition and were right-hand dominant. All subjects were native Danish speakers and had normal hearing.

2.2. Stimuli

Stimuli were the 8 syllables [b̥ɑ], [pʰɑ], [g̊ɑ], [kʰɑ], [b̥ɑː], [pʰɑː], [g̊ɑː] and [kʰɑː]. They were made from the exemplars [pʰɑː] and [kʰɑː] recorded by a 42 yo male. The fundamental voice pitch contour of these exemplars was flattened to 105 Hz with the PSOLA algorithm implemented in PRAAT [4]. Aspiration and the voiceless phase were removed from the exemplars to yield the syllables [b̥ɑː] and [g̊ɑː]. The vowel of these items was then truncated to provide syllables with short (120 ms) and long (200 ms) vowels. Linear gating was applied to the last 50 ms of all items. Table 1 shows the temporal specifics of the contrasts and when they differentiate according to a secondary sorting contrast.

Table 1: Mean times in ms (and range) from stimulus onset at which contrasts (rows) differentiate themselves according to a secondary contrast (columns).

	Length	Place
VOT	160 (120-199)	47 (15-79)
Length		132 (15-279)
Place		

2.3. EEG Recording and processing

A complete description of the experimental conditions and recording parameters is given in [5]. In short, there were 100 presentations of each syllable in an attend and a divert condition. In the attend condition subjects performed closed-set syllable identification, and in the divert condition they completed a visual discrimination task based on Kanji symbols. Low-density EEG recorded during both conditions underwent visual inspection and independent component analysis. Epochs were extracted between -200 and 400 ms relative to syllable onset, and baselined to the prestimulus data.

2.4. Accuracy correction and GFP contrast-attention differences for sorted features

Epochs from the attend condition, where the recorded response was either incorrect or absent, were omitted from further analysis. GFPs were then calculated for each syllable from the remaining epochs in both attention conditions. We applied subtraction formulas to the GFPs to calculate contrastive features according to a secondary sorting feature. For instance, to derive the VOT GFP differences, we applied formulas (1) and (2) so as to calculate for each place of articulation, and (3) and (4) so as to calculate for both VOTs.

$$(1)\quad \mathrm{VOT}_{\mathrm{Bilabial}} = \frac{\Sigma(\text{[pʰɑ] + [pʰɑː]})_i}{n} - \frac{\Sigma(\text{[b̥ɑ]+ [b̥ɑː]})_i}{n}$$

$$(2)\quad \mathrm{VOT}_{\mathrm{Velar}} = \frac{\Sigma(\text{[kʰɑ]+ [kʰɑː]})_i}{n} - \frac{\Sigma(\text{[g̊ɑ] + [g̊ɑː]})_i}{n}$$

$$(3)\quad \mathrm{VOT}_{\mathrm{Vowel\ short}} = \frac{\Sigma(\text{[pʰɑ]+ [kʰɑ]})_i}{n} - \frac{\Sigma(\text{[b̥ɑ] + [g̊ɑ]})_i}{n}$$

$$(4)\quad \mathrm{VOT}_{\mathrm{Vowel\ long}} = \frac{\Sigma(\text{[pʰɑː]+ [kʰɑː]})_i}{n} - \frac{\Sigma(\text{[b̥ɑː] + [g̊ɑː]})_i}{n}$$

The contrast differences from the attend condition were then subtracted from the divert condition to yield the timeseries given in Figs. 1-3. We analyzed the difference GFP timeseries by finding the maximum area between the curves in the poststimulus window (grey shaded area in Figs. 1-3). The time at which the curves intersect immediately prior to the maximum difference is interpreted as an initial separation in the processing of the attention contrast according to the two levels over which it is calculated.

3. RESULTS

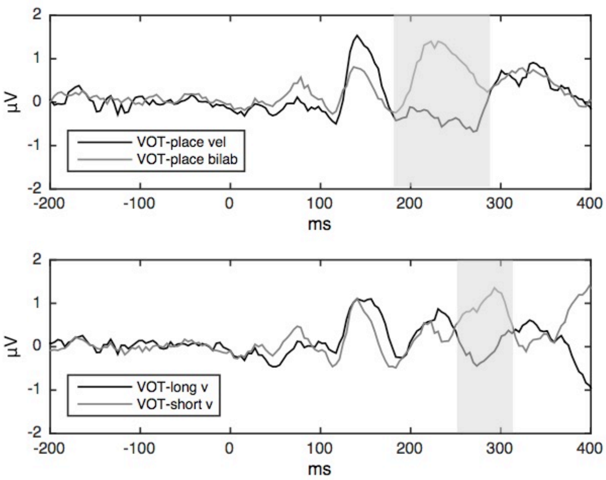
3.1. Secondary GFP contrasts

3.1.1. VOT

Fig. 1 shows local maxima in the VOT attention-contrast differences, when sorted for both place of articulation and vowel length, that occur at 140-150 ms poststimulus. This maximum is not evident in the other attention-contrast differences (Figs. 2 and 3) and is therefore probably linked to attentional processing of the VOT contrast. From the stimuli (see table 1), the temporal midpoint of the VOT contrast is auditorily available at a mean time of 47 ms after syllable onset, indicating that the attentional processing time involved in resolving this contrast is conservatively estimated at approximately 90 ms.

Within this initial peak it can also be seen that VOT attention-differences for place of articulation have different peak amplitudes which for velars are 1.11 μV (at 147 ms) and for bilabials are 0.78 μV (at 150 ms). In contrast to this, the VOT-vowel length differences are of equal amplitude, while the offset of the long vowel secondary contrast lags that of the short vowel by approximately 15 ms. This suggests that there is an integration of contrast processing whereby the VOT is processed in parallel with contrasts that occur earlier (place of articulation) and later (vowel length) in the stimuli.

Figure 1: VOT contrast-attention differences calculated according to place (upper) and vowel length (lower). In all figures, the shaded area indicates the maximum difference between the curves in the poststimulus window.

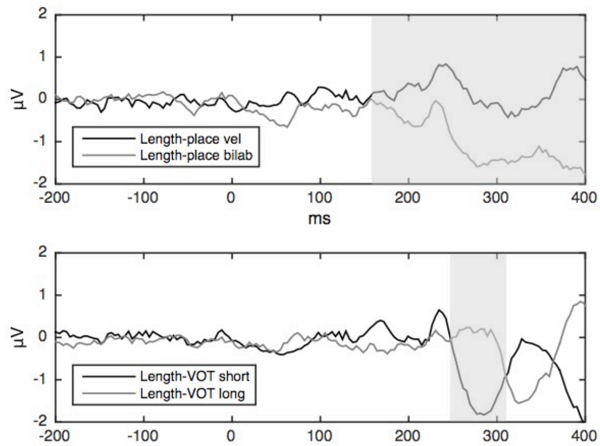


The local maxima subsequent to the VOT processing peak show a differentiation according to place of articulation in the 179-289 msec poststimulus window, during which the bilabial difference increases while the velar difference is between -0.1 and -0.7 μV. The VOT-vowel length maximum difference is between 250-314 ms poststimulus, and within this time window it can be seen that the long and short vowel attention differences are in inverse phase to each other.

3.1.2. Vowel length

The vowel length contrast differences, when calculated according to VOT, show abrupt negative deflections for the short VOT at 220 ms and the long VOT at 300 ms poststimulus (Fig. 2). The amplitude of both of these negative deflections is approximately 2 μV, which is the largest abrupt voltage change observed in all of the secondary contrasts. The temporal difference between the two levels of this secondary contrast suggests that the activity associated with vowel length is contingent upon VOT and is in a linear temporal relationship to this contrast. This in turn suggests that attentional tracking of the high amplitude vowels decreases according to the contrastive length of the VOT. When the vowel length contrast is calculated according to place of articulation there is a gradual differentiation between the velar and bilabial attention differences that begins at 152 ms poststimulus and continues to the end of the epoch.

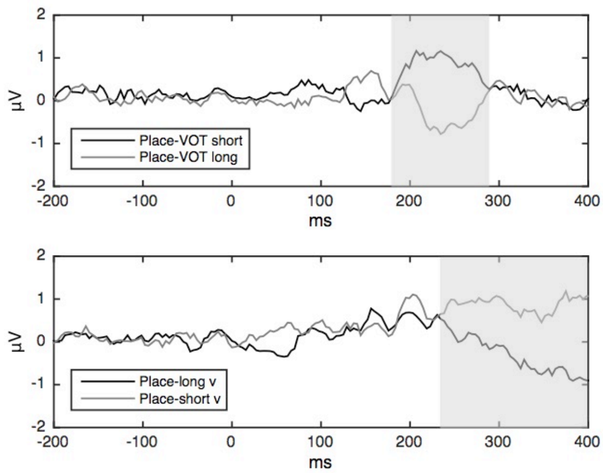
Figure 2: Vowel length contrast-attention differences calculated according to place of articulation (upper) and VOT (lower).



3.1.3. Place of articulation

Similar to the vowel length-place of articulation attention differences, the place-vowel length differences, differentiation continues until the end of the epoch from an initial divergence at 226 ms. Place of articulation calculated according to VOT shows a maximum differentiation in the 179-289 ms poststimulus window, where the secondary contrast differences are in opposite phase.

Figure 3: Place of articulation contrast-attention differences calculated according to place (upper) and VOT (lower).



3.2. Temporal relationships between stimuli and secondary contrasts.

Comparison of the beginning of the maximum area between levels of the secondary contrast (leftmost point of the shaded region in Figs. 1-3), and the difference between this time and the mean time after which stimuli differentiate themselves according to the contrasts are given in table 2. This shows that both VOT-place and its inverse diverge 132 ms after the temporal midpoint of the contrasts, as rendered in the stimuli. VOT-length and its inverse also diverge at the same time which was 90 ms after the temporal midpoint of the contrasts. However, the Length-place and the Place-length attention-contrast differences do not diverge at the same time, suggesting that these secondary contrasts are processed differently, or that this derived measure of differentiation does not correspond to attended perceptual processing of the Place-length and Length-place contrasts.

Table 2: Mean times in ms of the initial point of maximum differentiation and the difference between this point and the mean time from stimulus onset at which contrasts are available in the syllables.

Secondary contrast	Initial max (ms)	Initial diff. – stim (ms)
VOT-place	179	132
VOT-length	250	90
Length-place	152	30
Length-VOT	250	90
Place-VOT	179	132
Place-Length	226	94

4. DISCUSSION

These derived group neural timeseries suggest that there is an integration in the attentional processing of secondary contrasts, as for each primary contrast they differentiate maximally within overlapping poststimulus time windows. All contrasts show a period of maximum separation between the two levels over which they are sorted, that begins no later than 132 ms after the mean poststimulus time at which the contrast is auditorily available in the stimuli. The VOT attention-difference GFPs are of different phases in the period of maximum separation. The initial point of divergence prior to maximum separation is the same for both VOT-place and place-VOT, and VOT-vowel length and vowel length-VOT.

In contrast to the difference in phase observed in the attention-difference GFPs involving VOT, length-place and place-length timeseries show a gradual divergence of the attention-contrast differences that start later (179 or 226 ms) and continue to the end of the analysis epoch (400 ms). This may reflect continued repair processing related to perceptual uncertainty stemming from the low acoustic salience of the place of articulation cue. This explanation is plausible as it is generally accepted that place of articulation is a contrast, the accurate perception of which decreases rapidly with the addition of noise [9]. A related explanation for the late and long difference in the attentional processing observed in the place-length and length-place timeseries may be related to the duration of the cues. The stop difference between the velar and bilabial place of production occurred in the first 15 ms of the stimuli, whereas the VOT and vowel length contrasts both differed by approximately 80 ms.

When VOT was a primary contrast, there was an initial peak at 140 ms in both secondary contrasts, indicating that this is related to the resolution of the VOT contrast. This suggests that there is early attentional processing devoted to VOT, as this peak occurs approximately 90 ms after the temporal midpoint of the contrast as rendered in the auditory stimuli. The integration of the processing of secondary contrasts is also observed in the initial peak of the VOT contrasts as an attention-amplitude difference for place of articulation and a slight lag in the peak offset for vowel length. Subsequent local maxima in both the VOT-place of articulation and VOT-vowel length attention-differences may be related to what has previously been described as the 'double-on' neuronal response to VOT stimuli [6,7]. These results confirm that the dimensions of the double-on, particularly the second peak, is related to the attention of the listener, and, as reported in [8], contingent on other contrastive features, like place of articulation.

Transformed behavioral results from the same subjects may have shed light on perceptual dimensions of the contrasts that underlie these electrophysiology results and better substantiate the inferences that we draw. However, behavioral results would not be directly applicable to these electrophysiology results, as those from the attend condition were corrected for accuracy, a correction that was necessary in order to ensure precision in the analysis of differentiation between the two levels of the secondary contrasts. Despite the absence of applicable behavioral measures there is consistency in these results that is seen in the temporal correspondence between secondary VOT contrasts and their inverse. Similar subtraction methods may prove useful in probing the variation of attention that occurs with other contrastive speech features.

5. REFERENCES

[1] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432-434.

[2] Shtyrov, Y. 2010. Automaticity and attentional control in spoken language processing: Neurophysiological evidence. *The Mental Lexicon* 5, 225-276.

[3] Skrandies, W. 1990. Global field power and topographic similarity. *Brain Topography* 3, 137-141.

[4] Boersma, P., Weenink, D. 2018. Praat doing phonetics by computer [program]. V. 6.0.43 from http://www.praat.org/

[5] Morris, D. J., Tøndering, J., & Lindgren, M. 2019. Electrophysiological and behavioral measures of some speech contrasts in varied attention and noise. *Hearing Research* 373, 1-9.

[6] Steinschneider, M., Schroeder, C. E., Arezzo, J. C., Vaughan, H. G. 1994. Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 92, 30-43.

[7] Sharma, A., Dorman, M. F. 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 106, 1078–1083.

[8] Sharma, A., Marsh, C. M., Dorman, M. F. 2000. Relationship between N1 evoked potential morphology and the perception of voicing. *J. Acoust. Soc. Am.* 108, 3030–3035.

[9] Miller, G. A., Nicely, P. E. 1955. An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338–352.